Deploying Federal Geospatial Services in the Cloud:

Federal Geographic Data Committee (FGDC) and GSA GeoCloud Sandbox Initiative

Doug Nebert
USGS/FGDC
December 2010

Background on GeoCloud Sandbox Initiative

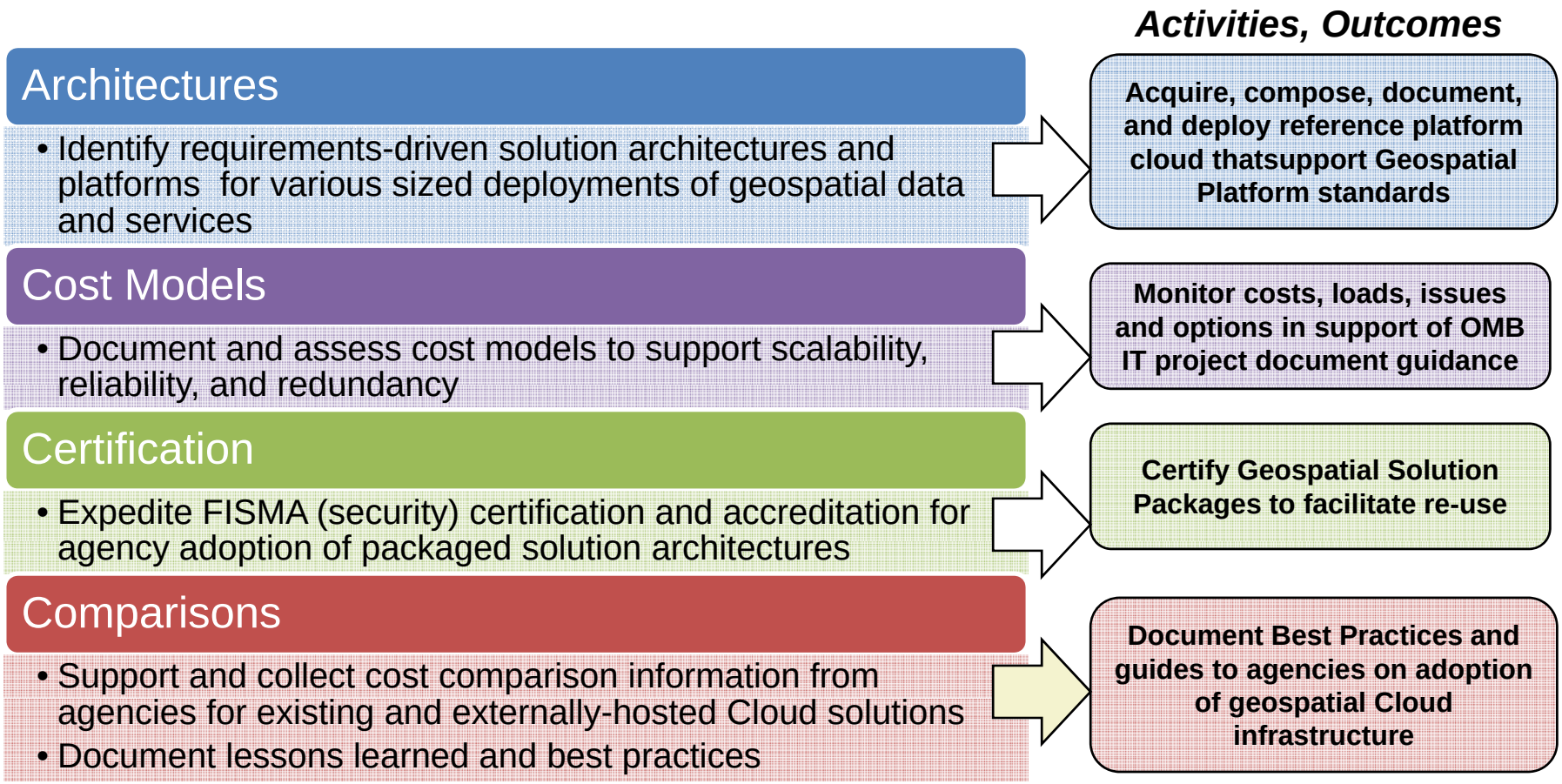
- Initiated as an Architecture and Technology Working Group activity in December 2009 with a call to federal agencies to nominate geospatial applications for testing in the Cloud environment for a one-year prototyping process coordinated by FGDC and GSA

- Eleven projects have been nominated by federal agencies as existing projects with existing software suites to be deployed in commercial- or government-hosted Cloud environments

- Two deployment environments (platforms) were abstracted from the nominated projects: Open source service stack on Linux64 and a Commercial service stack on Windows 2008 Server

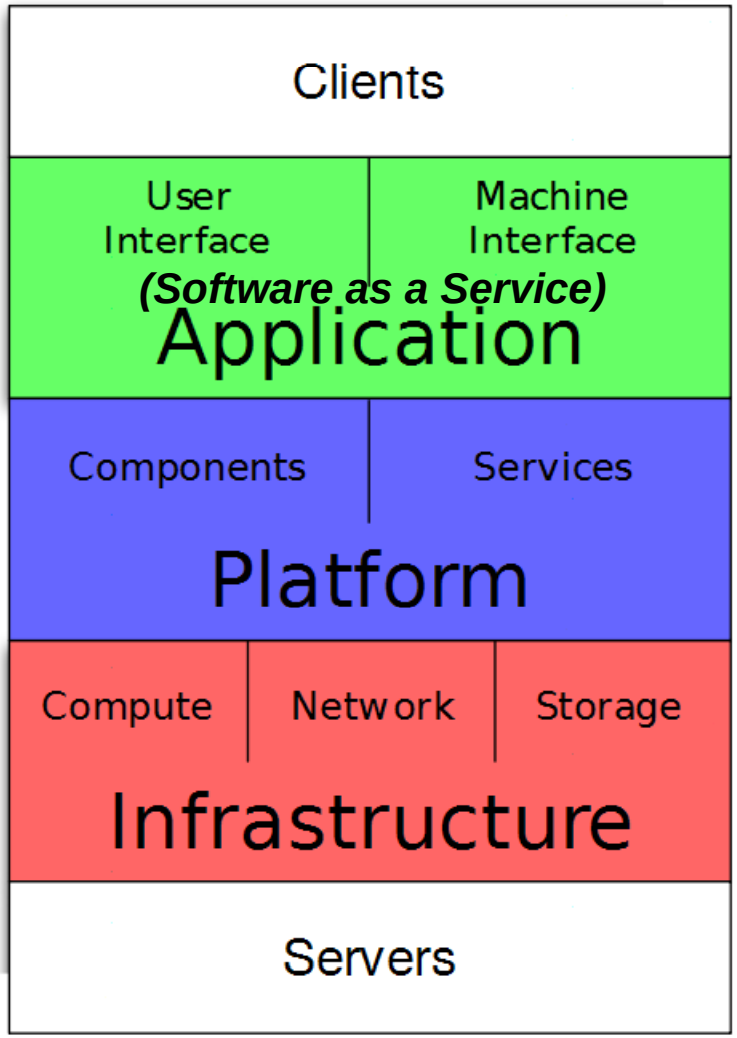
Deploying Federal Geospatial Services in the Cloud
GeoCloud Goals and Activities

One-year+ project to test and monitor externally-hosted Cloud data and service solutions for the geospatial domain, to support the Geospatial Platform activity



Notional Cloud Computing Stack

- **GSA apps.gov is now offering Infrastructure as a Service (IaaS) solutions for acquisition**
- **This is roughly equivalent to "shared-hosting" of raw computers with an operating system in the Cloud domain**
- **Platform as a Service (PaaS) provides configurable software components and services**



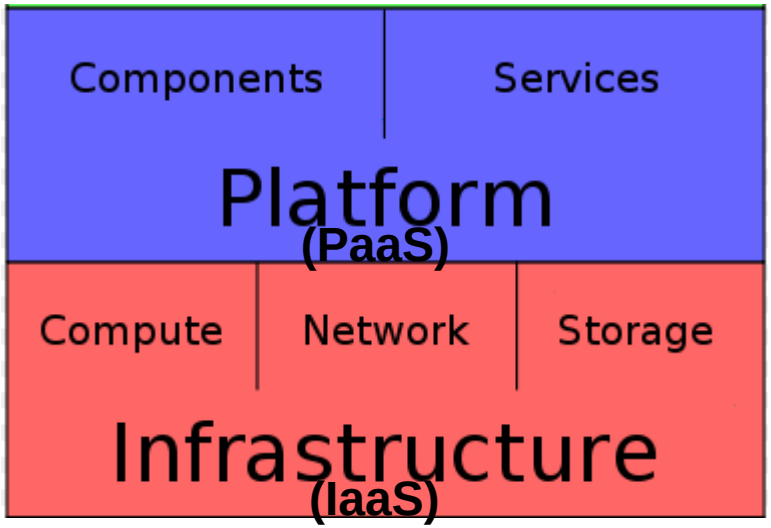
Cloud Computing Stack

http://en.wikipedia.org/wiki/Cloud_computing

Platform as a Service (PaaS)

"A *cloud platform* (PaaS) delivers a computing platform and/or solution stack as a service, generally consuming *cloud infrastructure* and supporting *cloud applications*. It facilitates deployment of applications without the cost and complexity of buying and managing the underlying hardware and software layers."*

The GeoCloud is piloting the deployment of candidate services via ***solution architectures*** (suites of software) by composing reference platforms *on top of* the GSA IaaS to provide common geospatial capabilities.



*** http://en.wikipedia.org/wiki/Cloud_computing**

GeoCloud Community Platform Value Proposition

Pool of geospatial projects identified for cloud migration

FGDC App 1

FGDC App 2

8 More …

Apps Need Platforms

Federal Platform Reference Implementations

GeoCloudCommunity Platform

Platform Savings

Platform building time & effort	Reduced maintenance costs	Faster deployment	Cost effective development / test

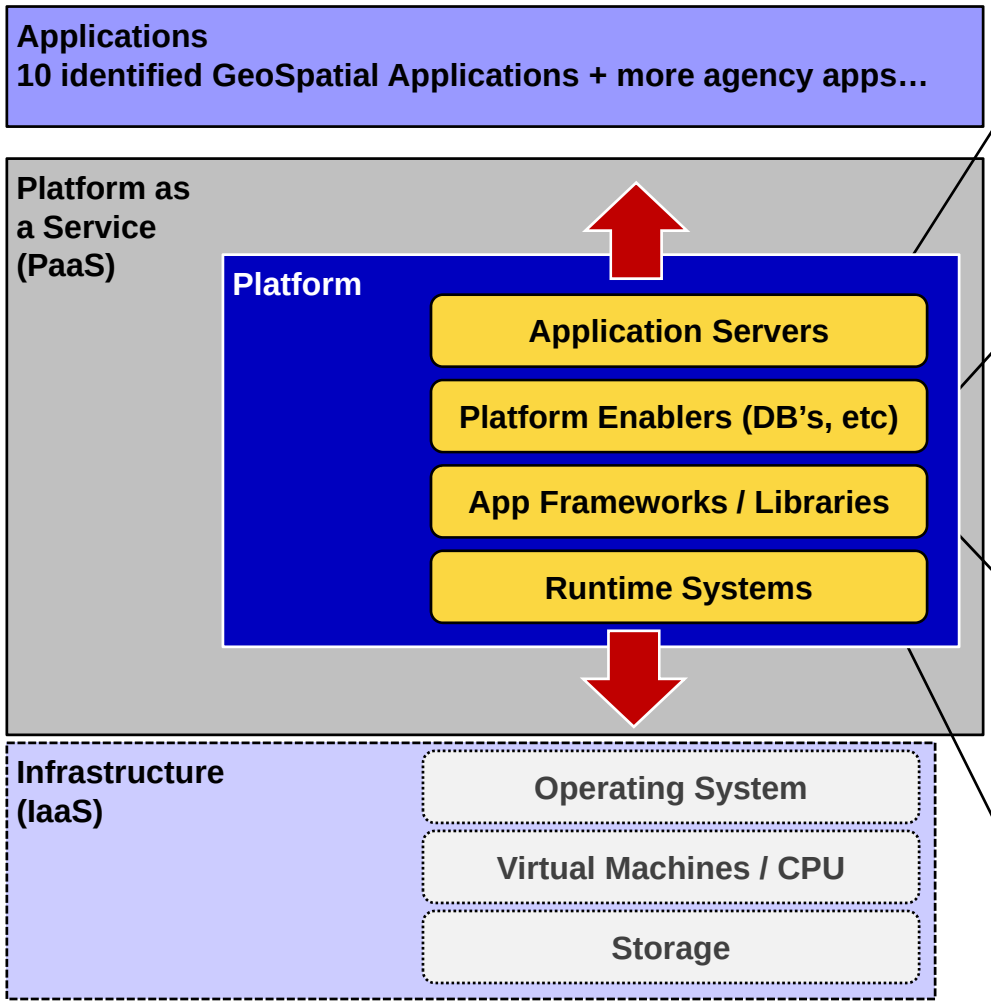
Platforms Amplify IaaS Savings

Infrastructure as A Service Savings

Hardware Savings	Operations Savings	Scalability Savings

Platform as a Service Architectural Framework

Distinguishing Application, Platform and Infrastructure



Application Servers: Provide the deployment environment for actual business applications with access to enablers, frameworks and runtimes, including interfaces and payload standards endorsed by the FGDC.

Platform Enablers: Platform enablers provide core supporting capabilities for developing , testing and deploying code, including DBMS, Directory, Queue, Service Bus, etc. A relational database is the most common enabler example but is not present in all platforms

Frameworks provide bundled access to common behaviors and services, which applications can rely upon.

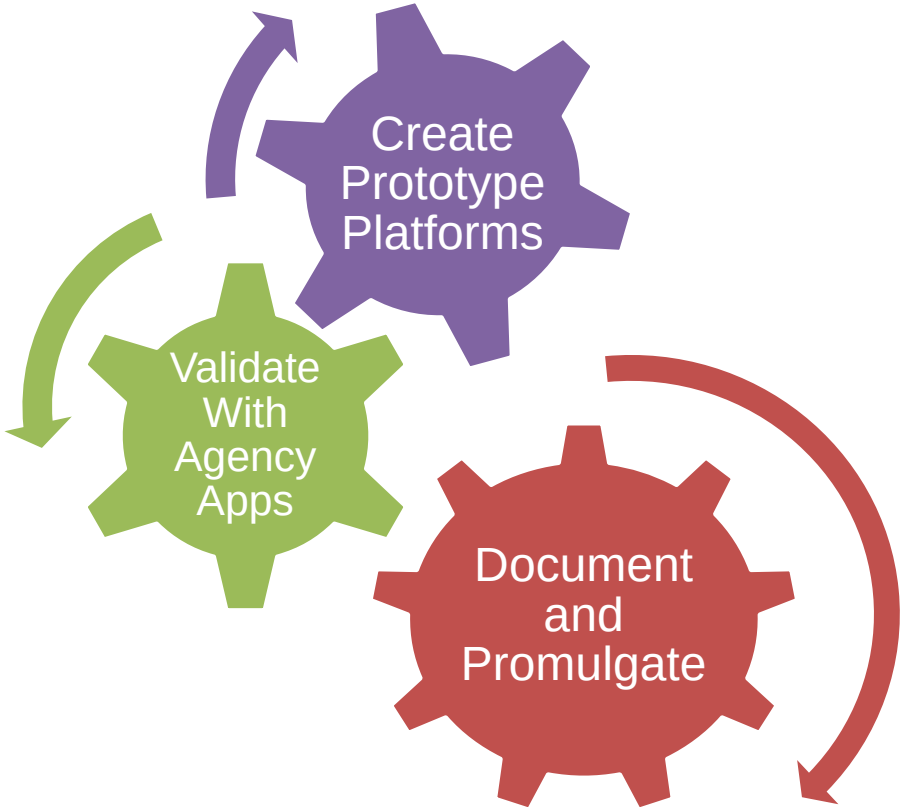
Libraries are reusable code modules which can be called directly from an application.

Frameworks and Libraries save time and expense freeing developers from having to build common code and behaviors.

Runtimes: provide the execution support for developing and running the code. Examples include Java, Python, Microsoft Common Language Runtime, etc.

Cloud Community Platform Service Activities

Primary Activity Cycle



Prototype

- Create series of Federal Cloud Computing reference platform prototypes
- Support the wide range of target apps

Validate

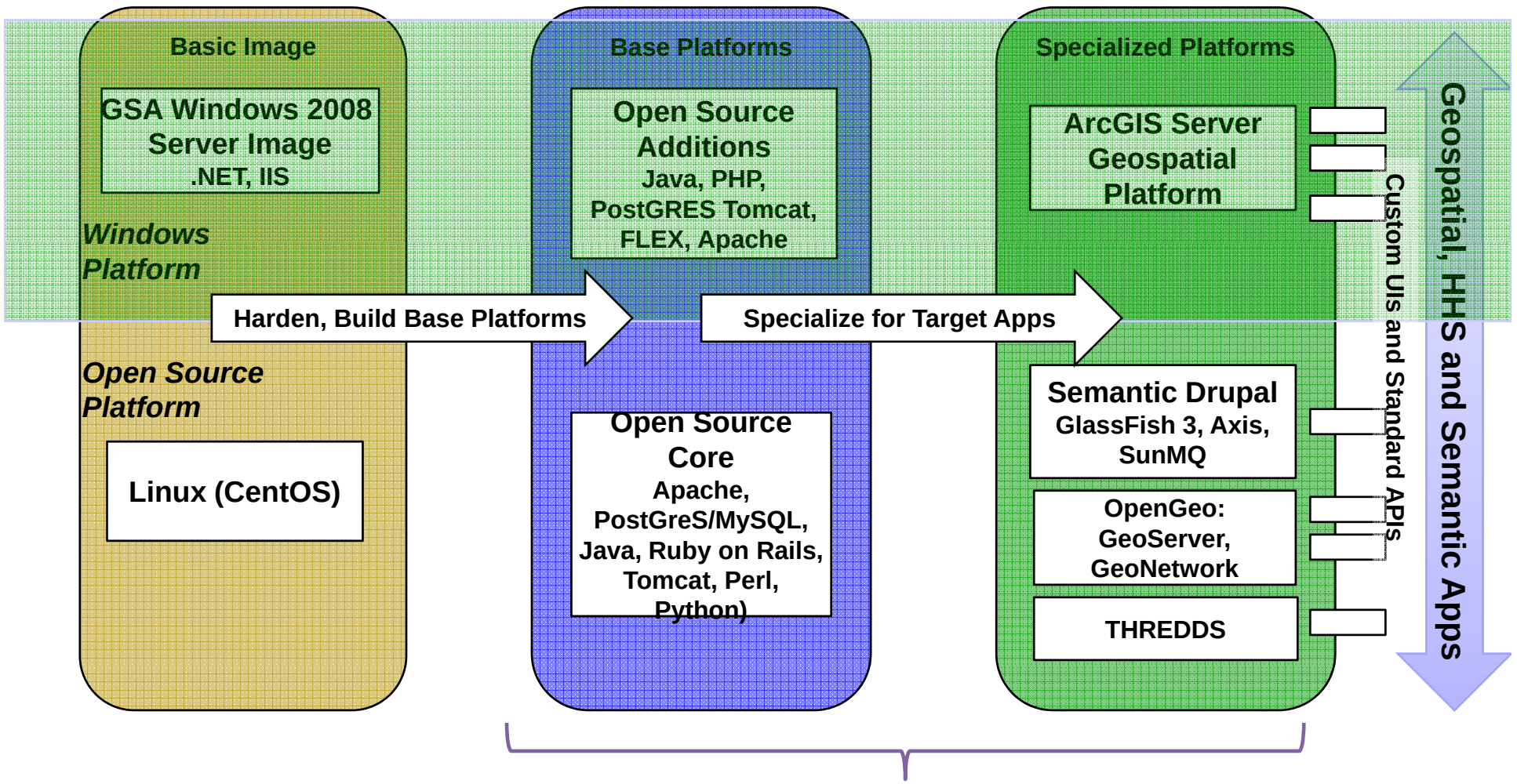
- Ensure platforms work with identified agency applications

Document

- Document best practices and lessons learned
- Document cost models and benchmarks
- Bundle the platforms and artifacts
- Promulgate to Federal Cloud Community of Interest

Cloud Community Platform Service Activities

Preliminary Platform Architecture



Note: Software requirements for the candidate platforms were dictated by the eleven projects nominated by federal agencies. Standards are dictated by the FGDC endorsed standards list, Commonalities were identified in defining the stack. Additional software suites could be defined in the future using the same design process.

Deployment options within GeoCloud

•Cost evaluation for each of the initial projects was performed using online ‘calculator’ based on data transfer, storage, CPU, and demand requirements

- Most projects could be feasibly hosted in AWS (~$350-500/month)
- Some projects were cost-prohibitive in the Cloud due to large data storage or transfer costs

•Amazon Web Services (EC2) was selected as the primary ‘public’ cloud computing environment for various sizes and numbers of virtual machines (AWS via Apptis is a GSA *apps.gov*IaaS offering on BPA)
•Dell/VMWarevCloud environment was selected for government-hosted cloud infrastructure – at USGS EROS Data Center

Project Synopsis

NWI Wetlands Mapper	AWS, Windows 2008, ArcGIS Server
US Census TIGER/Line Downloads	AWS, Linux64 (CentOS)
Integrated Ocean Observing System Catalog and Viewer (NOAA)	AWS, Linux64 (CentOS), GeoNetwork
NOAA ERDDAP	AWS, Linux64 (CentOS), THREDDS
EPA Lakes and Ponds	AWS, Windows 2008, ArcGIS Server

Project Synopsis, Continued

USGS National Elevation Dataset (NED) service and data	VMWarevCloud on Dell HW, Windows 2008, ArcGIS Server
USGS National Map Map Tile Cache	VMWarevCloud on Dell HW, Windows 2008 or Linux64
Particles in the Cloud (NOAA) particle tracking computational service for air or water dispersion/ diffusion	AWS, Linux64 (CentOS)
GEOSS Clearinghouse Catalog	AWS, Linux64 (CentOS)
USDA FSA or NRCS data service application	AWS, Windows 2008, ArcGIS Server

GeoCloud Status and Plans, as of January 2011

- Provided login credentials(encrypted key pairs) for AWS and its management interface to individuals using AWS
- Initiating image builds of OS plus base software dependencies for CentOS (RedHat) and scripting the geospatial module activation
- Scripting Windows 2008 Server using hardening guidelines used by GSA and other agencies with ESRI ArcGIS Server as image baseline
- Installation and monitoring of the deployment environment with agency-facing Web services (January-March)
- Initial public-facing Web Services (January)

Questions & Answers

- Contact Information:
  - Doug Nebert (ddnebert@fgdc.gov)
  - Robert Patt-Corner, eGlobalTech/GSA (robert.patt-corner@gsa.gov)